

Externalism and Incomplete Understanding

Philosophical Quarterly 54, pp. 287-294, 2004

Åsa Maria Wikforss
Stockholm University
asa.wikforss@philosophy.su.se

Externalism and Incomplete Understanding

By Åsa Maria Wikforss

1. Introduction

Social externalism, it is widely held, depends on the assumption that an individual may think with a concept despite having an incomplete understanding of this very concept. For instance, in Burge's famous arthritis thought experiment, the individual is said to have an incomplete grasp of the concept of arthritis and yet he is to be ascribed thoughts containing this concept.¹ The assumption of incomplete understanding, however, has made the externalist vulnerable to a certain type of individualist attack. How could one think with a concept one incompletely understands? It is one thing to say that we use *words* we understand incompletely, but the notion of partial understanding of one's own concepts is harder to make sense of. It also poses notorious difficulties when it comes to accounting for our reasoning abilities and actions.²

In her recent paper "Conceptual Errors and Social Externalism", Sarah Sawyer challenges the idea that social externalism depends on the assumption of incomplete understanding.³ Sawyer is responding to my paper "Social Externalism and Conceptual errors".⁴ In that paper I argued that Burge's famous thought experiment concerning the concept of arthritis relies essentially on the assumption that the individual, Bert, has an incomplete grasp of the concept of arthritis and that Bert

¹ T. Burge, 'Individualism and the Mental', *Midwest Studies* IV (1979), pp. 73-121.

² For a recent discussion of these difficulties, see J. Brown, 'Critical Reasoning, Understanding, and Self-Knowledge', *Philosophy and Phenomenological Research*, LXI (2000), pp. 659-676.

³ S. Sawyer, 'Conceptual Errors and Social Externalism', *The Philosophical Quarterly*, 53 (2003), pp. 265-273.

⁴ Å. Wikforss, 'Social Externalism and Conceptual Errors', *The Philosophical Quarterly*, 51 (2001), pp. 217-231.

makes a conceptual error when he utters “I have arthritis in my thigh”. I argued that this assumption should be questioned, and that once that is done we are free to grant Burge that Bert has the standard concept of arthritis, and yet the externalist conclusions will fail to go through. Sawyer agrees that Burge’s early thought experiment concerning the concept of arthritis appears to rely on the assumption of incomplete understanding, but she argues that this is just a special feature of that particular thought experiment and does not have any implications for social externalism generally. To illustrate this, Sawyer considers Burge’s later thought experiment concerning the concept sofa.⁵ This thought experiment, Sawyer argues, shows that the unifying principle behind social externalism is not the assumption of incomplete understanding, but just the idea that “referents themselves play a role in the individuation of concepts”.⁶

If Sawyer is right in this it is very interesting, since it would deprive the individualist of a major complaint against social externalism. In this note, however, I argue that Sawyer fails to establish that social externalism need not rely on the assumption of incomplete understanding. Burge’s sofa-thought experiment, even according to Burge himself, does not support social externalism but a form of physical externalism. Sawyer shows some awareness of this, but suggests that it does not matter much whether we call Burge’s later type of externalism ‘social externalism’ or ‘broad physical externalism’. However, I argue, more is at stake here than Sawyer suggests. If Burge’s later externalism is not a form of social externalism, Sawyer has failed to show what she set out to show. Moreover, precisely because Burge’s later externalism is not a type of social externalism it is very problematic, since it is far from clear that physical externalism can be extended, in the way Burge assumes, to

⁵ T. Burge, ‘Intellectual Norms and the Foundations of Mind’, *Journal of Philosophy*, 83 (1986), pp.

apply to all types of concepts. Sawyer's idea is that the unifying principle shows how this can be done, but I argue that Sawyer's principle cannot be used for that purpose. Further considerations are required to support the conclusions of Burge's later thought experiment and these considerations, I argue, do commit the externalist to the assumption of incomplete understanding.

2. Sawyer's argument

It is quite clear that Burge himself took his original thought experiment to rely on the assumption of incomplete understanding. The difference in the meaning of 'arthritis' in the two communities, Burge suggests, is a result of a difference in linguistic conventions: In the actual community, 'arthritis' is defined as a rheumatoid disease of the joints only, whereas in the counterfactual community 'arthritis' is defined more widely, as applying to rheumatoid diseases of the ligaments as well as of the joints.⁷ When, therefore, Bert utters 'I have arthritis in my thigh' he is not making an ordinary empirical error, according to Burge, but a conceptual one. His utterance betrays an incomplete understanding of the meaning of 'arthritis', of the standard concept of arthritis.⁸ When Bert's twin utters the same words, however, the difference in community conventions implies that he has made a true utterance. Hence, Bert and his twin express different thoughts, simply as a result of relying on different linguistic conventions.

Now, Sawyer is aware of this, and quotes several passages of Burge's where he says that the thought experiment depends on the assumption of incomplete understanding. However, she argues that it does not follow that social externalism

697-720, at pp. 707-708.

⁶ Sawyer (2003), p. 272.

⁷ Burge (1979), p. 78.

⁸ Ibid. pp. 82-83.

generally relies on the assumption of incomplete understanding. To make her case, Sawyer refers to Burge's later thought experiment concerning the concept of sofa.⁹ In that experiment, recall, we are to imagine that A in the actual world, and B in a counterfactual world, both use the word 'sofa' competently. However, they proceed to develop non-standard theories about the objects in their environment called 'sofas', and start to doubt the truth of the statement "Sofas are furnishings to be sat upon". A's doubt, as it turns out, proves unfounded: It is indeed part of the nature of sofas, in the actual world, that they are furnishings to be sat upon. B's doubts, however, prove to be correct: The objects that B is confronted with look like sofas but function as works of art and would collapse under a person's weight. This implies, according to Burge, that "there are no sofas in B's situation, and the word form 'sofa' does not mean sofa."¹⁰ Despite being physically identical, A and B mean different things by 'sofa' and have different 'sofa'-concepts.

The important difference between this thought-experiment and the 'arthritis'-one, as Burge himself emphasizes, is that in this experiment there is no appeal to linguistic conventions and the speaker is not said to have an incomplete grasp of the conventional meaning of the term in question.¹¹ The difference in meaning derives not from any difference in conventions but from a difference in the nature of the objects called 'sofa' in A's and B's worlds. When, therefore, A thinks that sofas may not be furniture meant for sitting, he is not displaying an incomplete grasp of the conventional meaning of 'sofa', but is simply hypothesizing a non-standard theory about sofas. Sawyer takes this to show that social externalism does not essentially depend on the assumption of incomplete understanding. The crux of externalism, she argues, lies not with the assumption of incomplete understanding but with the idea

⁹ Burge (1986).

that “referents themselves play a role in the individuation of concepts”. This idea, she suggests, “is the principle that unifies the thought experiments”.¹² The reason ‘sofa’ has a different meaning, expresses a different concept, in B’s world than in A’s is simply that ‘sofa’ has a different extension in the two worlds.

Let us grant, for the moment, that Burge’s sofa-thought experiment does not rely on the assumption of incomplete understanding. The question is how this could possibly refute the thesis that social externalism depends on this assumption. After all, the sofa-thought experiment does not appeal to the *social* environment at all, to the conventions or practices of the linguistic community, but to the *physical* one; i.e. to the nature of the objects called ‘sofa’ in the two worlds. Indeed, Burge himself is very explicit on this point, and argues that what the thought experiment shows is that “even where social practices are deeply involved in individuating mental states, they are often not the final arbiter.”¹³

Sawyer, in fact, concedes this point and notes that Burge himself did not seem to take the sofa-experiment to support social externalism.¹⁴ She suggests, however, that it is of little consequence whether the ensuing externalism can be properly described as social or not. According to Sawyer the value of Burge’s second thought experiment is that it shows that there is a type of physical externalism that is not limited to natural kind terms, but applies more broadly. It is therefore not important, she holds, whether the externalism in question is labeled ‘social externalism’ or ‘broad physical externalism’; either way the result is the same – a type of externalism that applies extremely widely, to any type of word or concept.

¹⁰ Ibid. p. 708.

¹¹ Ibid. p. 709.

¹² Sawyer (2003), p. 272.

¹³ Burge (1986), p. 707.

¹⁴ Sawyer (2003), p. 273.

However, this move is much too swift. If Burge's later thought experiment does not support social externalism, then Sawyer has failed to supply any reason to endorse the main thesis of her paper – i.e. the thesis that social externalism does not rely on the assumption of incomplete understanding. And while it obviously does not matter much what we *label* Burge's later externalism, it is clear that it is *not* a form of social externalism, even as defined by Sawyer herself at the outset: "The thesis of social externalism is the thesis that many of a subject's mental states and events are dependent for their individuation on the subject's social environment."¹⁵ By conceding that Burge's later thought experiment does not support social externalism, therefore, Sawyer concedes that she has failed to show what she set out to show.

This still leaves an interesting question, however, one that is the real topic of Sawyer's paper: Does Burge's later thought experiment show that there is a kind of *physical* externalism that applies just as broadly as social externalism, and that does not rely on the assumption of incomplete understanding? If this is so, the externalist could simply give up on social externalism, without much loss. To answer this question, we have to take a closer look at Sawyer's unifying principle.

3. A unifying principle?

Sawyer's suggestion, again, is that the unifying principle behind the thought experiments is the claim that "concepts are individuated partly by their referents rather than entirely by what the subject thinks is true of the referents". Now, this claim, as it stands, is rather vague and provides little more than a rough characterization of content externalism: Content externalism takes concepts to be individuated partly by their referents. But this does not tell us anything about *how*

¹⁵ Ibid. p. 265.

referents serve to individuate concepts, and it leaves it an open question whether this individuation entails that the individual has an incomplete grasp of concepts that are externally individuated. However, in her discussion Sawyer hints at a further, more precise, claim. She emphasizes that ‘sofa’ has a different extension in A’s and B’s worlds and indicates that this difference in extension generates the conclusion that the concepts of A and B differ.¹⁶ This suggests the following underlying principle: *A difference in reference (extension) implies a difference in concepts.*

Burge, in fact, appeals to this principle in defense of content externalism. For instance, the principle plays a central role in Burge’s argument for why Putnam’s meaning externalism should be extended to concepts and content. If ‘water’ has a different reference on Twin Earth than it does on Earth, Burge argues, then it must also express a different concept on Twin Earth.¹⁷ In a later paper Burge makes the underlying principle explicit and argues that it applies to a large number of what he calls ‘empirically applicable terms’ (nouns and verbs that apply to everyday, empirically discernable objects): “Although the reference of these words is not all there is to their semantics, their reference places a constraint on their meaning, or on what concept they express. In particular, any such word *w* has a different meaning (or expresses a different concept) from a given word *w*’ if their constant referents, or ranges of application, are different. That is part of what it is to be a non-indexical word of this type.”¹⁸

¹⁶ Ibid. pp. 271-72.

¹⁷ What Putnam fails to note, Burge argues, is that the difference in reference between ‘water’ on Earth and ‘water’ on Twin Earth, affects “oblique occurrences in ‘that’-clauses that provide the contents of their mental states and events.” (T. Burge, ‘Other Bodies’, in A. Woodfield *Thought and Object*, Oxford: Clarendon Press, (1982), pp. 97-120, at p. 101.)

¹⁸ Burge 1989, ‘Wherein is Language Social?’ In A. George, *Reflections on Chomsky*, (1989), pp. 175-191, at p. 181. Burge appeals to this principle already in ‘Individualism and the Mental’: “On any systematic theory, differences in the extension – the actual denotation, referent, or application – of counterpart expressions in that-clauses will be semantically represented, and will, in our terms, make for differences in content.” (1979, p. 75).

Sawyer can therefore find support for her principle in the writings of Burge. However, this does not help Sawyer, since the principle cannot do the work she wants it to do. Depending on how it is understood, the principle is either false or cannot be used without begging the central question. Let me explain.

First, in one sense the principle is obviously false. It is a trivial truth that our word 'sofa' could have a different extension in another possible world, without it following that the word has a different meaning in that world. For instance, there is a possible world in which all sofas are made of leather but this is still a world in which 'sofa' has the same meaning as in the actual world (or else the possible world could not be thus described). Consequently, the fact that 'sofa' has a different extension in B's world than in the actual world, does not yield the externalist conclusion that the word has a different meaning and expresses a different concept in B's world.¹⁹

It might be objected that this reflects a misunderstanding of the argument. The claim is not that our word 'sofa' has a different extension in B's world (the extension of poorly made sofas that break when sat upon) than in the actual world. Rather, the claim is that our word 'sofa' does not apply at all to the objects called 'sofa' in B's world, since *those objects are not sofas* (and vice versa, B's word 'sofa' does not apply to the objects called 'sofa' in our world). The objects in B's world are simply not within the extension of our word, and so B's word 'sofa' must have a different meaning. Thus, when A theorizes that sofas may not be furniture meant for sitting he is theorizing about *a different type of object* than B does when he develops his non-standard theory. Indeed, this is Sawyer's explicit reasoning at one point. The actual situation, she says, contains sofas whereas the counterfactual situation does not. Consequently, she argues, "A's community and B's community do not have different

theories about the same things, but have, rather, different theories about *different* things.”²⁰

Construed this way Sawyer’s principle is obviously correct and yields the desired conclusion. If there are no sofas in B’s world, then his term ‘sofa’ has a different meaning and expresses a different concept than our term ‘sofa’ does. However, it should be clear, this reasoning begs the central question. Namely, why should we accept the claim that the objects in B’s world are not sofas? That is, we can grant that *if* A’s term ‘sofa’ does not apply to the objects B calls ‘sofa’, then A and B have different ‘sofa’-concepts, but what we need is a reason to accept the antecedent in the first place – i.e. a reason to accept the claim that the objects in B’s world are not sofas. Sawyer’s principle simply has no bearing on this topic.

Sawyer appears to hold that Burge need not *argue* for the claim that there is a difference in reference but can simply *stipulate* it. As Sawyer puts it: “The thought experiment is Burge’s, and we surely cannot deny him this stipulation.”²¹ It is of course correct that Burge is free to make stipulations. However, if the difference in reference is merely stipulated Burge cannot, as he now does, appeal to our intuitions about the English language and his thought experiment could not provide evidence for a certain thesis about the semantics of our language.²² Similarly, Putnam could have stipulated that ‘water’ has a different reference on Twin Earth, rather than arguing for it, but had he done so the thought experiment could not be used the way Putnam uses it -- to motivate a certain account of the semantics of our term ‘water’-- and the

¹⁹ For a discussion of this see S. Häggqvist, *Thought Experiments in Philosophy*, Stockholm: Almqvist & Wiksell International, (1996), p.177.

²⁰ Sawyer (2003), p. 272.

²¹ Sawyer (2003), p. 271.

²² I discuss this point briefly in Wikforss (2001), p. 224. For an illuminating discussion of stipulations and thought experiments see Häggqvist (1996), pp. 146-147.

ensuing debate concerning whether ‘water’ does apply to XYZ would be utterly pointless.²³

Sawyer’s principle therefore cannot do the work she wants it to do. Construed the first way, the principle is false, since there are many possible worlds in which our term ‘sofa’ has a different extension without therefore expressing a different concept. Construed the second way, the principle is true, and can be used to motivate the move from reference externalism to content externalism, but not to motivate the reference externalism in the first place. To make a case for the claim that our term ‘sofa’ does not apply to the objects in B’s world, considerations of a quite different kind are required. The question, then, is whether these considerations will commit the externalist to the assumption of incomplete understanding. Although this question cannot be fully examined here, I will end by suggesting that there are strong reasons to believe that the further considerations required will indeed commit the externalist to the assumption of incomplete understanding.

4. Incomplete Understanding Again

It is no doubt true that many people share the intuition that ‘sofa’ does not apply to the objects in B’s world. The most straight-forward explanation of this intuition is that the belief that sofas are to be sat upon is so central to the meaning of our word ‘sofa’ that the objects in B’s world could not possibly fall within its extension, since they are so brittle that they cannot be sat upon. After all, terms for artifacts are typically given functional definitions, and it is not implausible that ‘sofa’ should be given such a definition. However, it is obvious that this reply is not available to Sawyer. If this is

²³ This debate has been ongoing since Putnam’s paper first appeared. See for instance D. H. Mellor, “Natural Kinds”, (1977). (Reprinted in *The Twin Earth Chronicles*, eds. A. Pessin and S. Goldberg, Armonk, NY: M.E. Sharpe.) Mellor writes: “I agree that ‘water’ had (tenselessly) the same extension in

the reason our term ‘sofa’ does not apply to the objects in B’s world, then it follows that A who doubts that sofas are furnishings to be sat upon *does* display an incomplete understanding of the concept of sofa. That is, the thought experiment would rely on the assumption of incomplete understanding after all.

The challenge Sawyer faces (along with Burge) is therefore to give an argument for why the objects in B’s world are not sofas, without appealing to intuitions about the conventional meaning of ‘sofa’. I think this is a formidable challenge, and that it is one reason Burge’s later externalism has achieved much less attention than his earlier social externalism. In the case of natural kind terms the challenge is met by appealing to an essentialist conception of natural kinds. On Putnam’s view natural kinds have an essential underlying micro-structural property, and when we use a term as a natural kind term, we intend to denote this underlying property, whether or not we have any knowledge of it. Thus, XYZ is not within the extension of our term ‘water’, even though it is not part of the *meaning* of our term that water is H₂O.²⁴ A similar move, it would seem, is required in the case of Burge’s later thought experiment. However, what sense are we to make of the idea that sofas, knives, clothing, rope, pottery, tables, watches, etc. (to mention a few of Burge’s examples) have an essence given by ‘nature itself’, independently of our classificatory practices; an essence that we intend to pick out when using terms like ‘sofa’, ‘knife’, ‘pottery’, etc.²⁵ It is one thing to endorse essentialism in the case of natural kinds, quite another to extend the essentialism to non-natural kinds, such as artifacts.

Let us assume, however, for the sake of argument, that the challenge can be met and that some form of generalized Aristotelian essentialism is adopted. That is,

1750 as it had in 1950; what I deny is that at either time that extension was different on Earth and on Twin Earth.” (p. 71)

²⁴ For a detailed discussion, and criticism, of the semantic assumptions underlying Putnam’s externalist account of natural kind terms see my paper ‘Naming Natural Kinds’ (forthcoming in *Synthese*).

let us assume that artifacts and other non-natural kinds have essences just like natural kinds, and that these essences serve to individuate the relevant concepts. Does it follow that the content externalist is freed from the assumption of incomplete understanding?

To answer this question, let us have a closer look at Burge's own discussion of the 'sofa'-thought experiment. As mentioned above, Burge makes it quite clear that A does not have an incomplete understanding of the conventional meaning of 'sofa'. However, it only follows from this that A does not have an incomplete understanding of the *concept* of sofa, if conventional meaning and concepts coincide. In Burge's original thought experiment, concepts and conventional meaning do so coincide. However, the point of Burge's thought experiment concerning the concept of sofa is precisely that conventional meaning and concepts do *not* coincide. The conventional meaning of a term, Burge suggests, is given by the community use, or more precisely, by the normative characterizations that the experts, upon reflection, have come to agree on. Such meaning characterizations, however, can be rationally doubted, and this shows, according to Burge, that there is another notion of meaning, that of 'cognitive value'.²⁶ This latter notion of meaning is determined by the 'real nature' of the objects referred to, rather than by expert use, and it is therefore something we can all be wrong about. Concepts, Burge argues furthermore, are tied to this second notion of meaning and not to the notion of conventional meaning. It is because concepts depend for their individuation on the actual nature of the objects referred to that the conventional definitions or conceptual explications we give may be rationally doubted and revised. For example, Dalton's definition of 'atom' in terms of indivisibility had

²⁵ See Burge (1986), p. 709.

²⁶ In Burge (1989), p. 181, he calls this other notion of meaning 'translational meaning' since, he suggests, it is what is preserved in translation. For the same distinction, see also Burge "Concepts, Definitions and Meaning", *Metaphilosophy*, 24, (1993), pp. 309-325.

to be revised since, as it turns out, atoms are in fact divisible. This revision does not show that Dalton had a different concept of atom than we do, Burge argues, but merely that his grasp of the concept of atom was less than complete.²⁷

Given this picture, it is clear that the fact that A has a complete understanding of the conventional meaning of ‘sofa’, does not tell us anything about his grasp of the *concept* of ‘sofa’. Moreover, it is quite clear that an essential presupposition of the thought experiment is that A *does* have an incomplete grasp of the concept of sofa. What A hypothesizes, again, is that sofas are not furniture meant for sitting. As it turns out, however, it is part of the real nature of the objects called ‘sofa’ in A’s world to be pieces of furniture meant for sitting. This, again, is the reason ‘sofa’ expresses a different concept in B’s world. Consequently, A’s doubts show that he has an incomplete understanding of the concept of sofa, just as much as Bert’s doubts show that he has an incomplete understanding of the concept of arthritis. In both cases the individual fails to grasp some central conceptual connections, only in A’s case this is a result not of a poor grasp of the linguistic conventions but of a poor grasp of the real definition of the objects referred to.

Burge’s later type of thought experiment, even by his own lights, is therefore committed to the assumption of incomplete understanding. This is not an accidental aspect of Burge’s experiment, but is a direct consequence of the suggestion that concepts are individuated by the real definitions of things: Since real definitions are distinct from epistemic or conventional definitions, any such suggestion will imply that speakers typically have an incomplete understanding of the very concepts that go

²⁷ Burge (1986), p. 718. For a further discussion of Burge’s later account of the relationship between conventional meaning and cognitive value, see my paper ‘An A Posteriori Conception of Analyticity?’, forthcoming in *Grazer Philosophische Studien*.

into their own thoughts and reasoning.²⁸ Indeed, the incompleteness in question is much more radical and pervasive than that implied by social externalism. On this view, not only non-experts but all members of the community may have an incomplete grasp of their own concepts, and the grasp cannot be improved upon without undertaking empirical investigations of the physical environment. The externalist who wishes to avoid the assumption of incomplete understanding, is therefore ill-advised to follow Burge and tie concept individuation to real definitions.

²⁸ The same holds for natural kind externalism once it is applied to concepts. If Oscar does not know that water is H₂O, then he has an incomplete grasp of the real definition of water (assuming essentialism) and thus of the concept of water. This is so even if it is not part of the conventional meaning of 'water' that water is H₂O.